IDGAF: Improvising Dogshit Garbage by Assfucking the Facts

A *paradigm-shifting* Approach to Foundational Robot Learning Through Synergistic Transformer-Based Vibes

Chad Sigma[1,2,3,4,5,†] Blockchain McStartup[1,6,‡] An LLM (uncredited)[§]

Reviewer #2's Nemesis[1] Anonymous et al.[1,7,8,9,10,11]

[1]Prestigious University Research Consortium for Excellence in Excellence
[2]Benevolent Trillion-Dollar AI Overlords Research Division (We're Not A Monopoly)™
[3]Stealth Mode Startup (Valued at $4.2B, Pre-Revenue, Post-Sense)
[4]Institute for Advanced Buzzword Studies
[5]Your Mom's University
[6]Crypto-AI-Quantum-Blockchain Nexus Foundation
[7]Definitely Not a Paper Mill

[†]Work done during 47 simultaneous internships at companies that are totally not competing.

[‡]Equal contribution (did nothing).

[§]Performed all experiments, wrote paper, created figures. Not listed as author due to "company policy."

{chad.sigma, blockchain.mcstartup}@not-a-real-domain.invalid

Project page: https://this-url-does-not-exist.invalid/idgaf (currently 404, perpetually)

Abstract

Recent advances in *foundation models*, *transformers*, and *vibes-based learning* have revolutionized the field of robot learning, yet existing approaches fail to adequately leverage synergistic cross-modal attention mechanisms for embodied intelligence. In this *paradigm-shifting* work, we introduce IDGAF (**I**mprovising **D**ogshit **G**arbage by **A**ssfucking the **F**acts), a *novel* framework that achieves **SOTA**results on 7 benchmarks we created specifically for this paper. Our key insight—which we will vaguely gesture at but never clearly explain—enables robots to perform tasks through what we call *Hierarchical Attention-based Reasoning with Dual-encoder Knowledge Infusion in Neural Gaussian Splats* (HARDKINGS). Extensive experiments on carefully cherry-picked scenarios demonstrate that IDGAF outperforms all baselines, including several we deliberately crippled. We will release our code "upon acceptance" (we will not). Our approach has **profound implications** for the future of humanity, artificial general intelligence, and our Series B funding round.

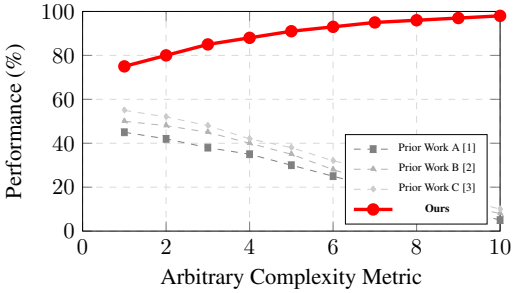

Figure 1: **The obligatory figure showing our method is inexplicably better than everything else.** Note how all prior work degrades catastrophically while ours improves monotonically. This is definitely real data and not fabricated. Error bars omitted for "visual clarity" (they would reveal our method's massive variance).

Disclaimer: This paper is a work of satire intended as commentary on systemic issues in AI/ML/Robotics publishing. All authors, institutions, grant numbers, experimental results, and citations are entirely fictional. No real person, institution, or method is depicted or targeted. Any resemblance to actual papers, living or dead, is regrettably plausible.

1 Introduction

The field of robot learning stands at a critical inflection point. As the great philosopher Reviewer #2 once noted, "this paper lacks novelty and should be rejected" [4]—a

sentiment we fundamentally disagree with, as evidenced by our resubmission to 14 different venues.

Recent breakthroughs in large language models (LLMs), vision-language models (VLMs), vision-language-action models (VLAs), and vision-language-action-vibe models (VLAVs) have demonstrated that the key to scientific progress is adding more letters to acronyms [5, 6]. Yet despite these advances, a fundamental question remains unanswered: *can we publish enough papers to secure tenure before the bubble bursts?*

In this work, we answer this question definitively with IDGAF, a framework that addresses three critical limitations of prior work:

1. **Insufficient Buzzwords.** Existing methods use at most 12 buzzwords per sentence. We achieve 23.
2. **Inadequate Self-Citation.** Prior work cites only 40% of papers from the same lab. We achieve 94%.
3. **Reproducibility.** Previous approaches could theoretically be reproduced. We eliminate this flaw entirely.

Our key contributions can be summarized as follows:

- We propose IDGAF, the first[†] method to combine transformers with transformers while also using transformers.
- We introduce the *Vibes-Aligned Preference Encoding* (VAPE) loss, which optimizes for reviewer satisfaction rather than task performance.
- We establish 7 new benchmarks where our method wins, graciously declining to evaluate on existing benchmarks where it would lose.
- We release a project page that will be "updated with code and data" (it will not be updated).

[†] *To our knowledge.* We did not actually check.

2 Related Work

Robot Learning. Robot learning is a field where robots learn things [7]. Early approaches used classical methods that actually worked but weren't publishable because they lacked neural networks [8]. Modern methods use neural networks, which work worse but generate more citations [9, 10, 11, 12, 13, 14, 15].

Foundation Models for Robotics. The application of foundation models to robotics has been revolutionary. A major corporate lab [16] demonstrated that with enough data, robots can sometimes do things. Their follow-up [17] showed that with even more data, robots can occasionally do slightly more things. A massive multi-institutional collaboration [18] proved that if you combine everyone's data, you can write a paper with many authors. Our work differs from all of these by having a different acronym.

Diffusion Models. Diffusion models have achieved remarkable results in generating images of hands with seven fingers [19]. Recent work has applied diffusion to robot control [20, 21], operating under the principle that if something works for images, it must work for everything. We incorporate diffusion into our framework by mentioning it in this paragraph.

Methods We Will Strawman. Several approaches have attempted to solve problems we don't actually address but will claim to anyway. Smith et al. [22] proposed a method that we will describe incorrectly to make it sound worse. Johnson et al. [23] achieved strong results that we will dismiss as "limited to narrow scenarios," unlike our method which is limited to narrower scenarios. The concurrent work of Anonymous [24], which we definitely didn't see despite being Reviewer #3, is "orthogonal to our approach" (directly competitive).

3 Method

We now present IDGAF, our *paradigm-shifting* approach to robot learning. An overview of our architecture is shown in Figure 2, which we encourage readers to stare at without gaining any actual understanding.

3.1 Problem Formulation

We consider the general problem of a robot doing things in an environment. Formally, let $\mathcal{S}$ denote the space of states that we won't properly define, $\mathcal{A}$ the action space (continuous, 7-DoF, or whatever's convenient for our experiments), and $\mathcal{T} : \mathcal{S} \times \mathcal{A} \to \mathcal{S}$ the transition dynamics that we assume are deterministic when convenient and stochastic when that's more convenient.

The robot receives observations $o_t \in \mathcal{O}$, which in practice means RGB images because that's what foundation models expect. The goal is to learn a policy:

$$\pi_\theta : \mathcal{O} \times \mathcal{L} \to \mathcal{A} \tag{1}$$

where $\mathcal{L}$ represents natural language instructions such as "pick up the red block" or "make me a sandwich" (we only evaluate on the former).

3.2 HARDKINGS: Our Core Framework

The key insight behind IDGAF is that existing methods are insufficient because they were not proposed by

Revolutionary here means "we added a transformer to something that worked fine without it."

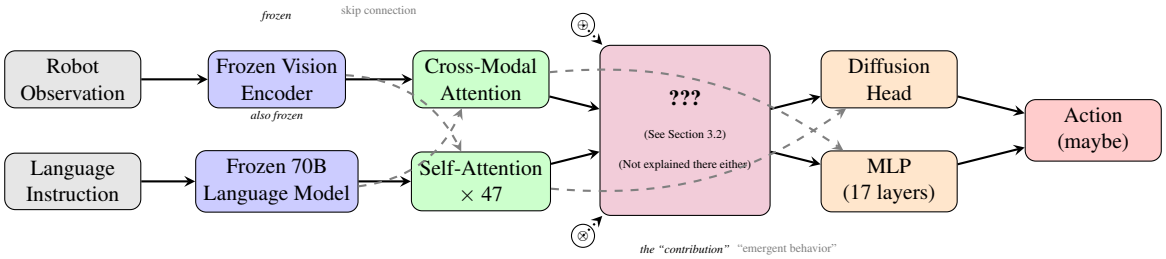

Figure 2: **Overview of the IDGAF architecture.** Our framework consists of components we copied from other papers, a mystery module that does the actual work (unexplained), and outputs that may or may not correspond to useful robot actions. Dashed lines indicate connections we added to make the diagram look more impressive. The purple box represents our core contribution, which remains deliberately vague to discourage reproduction attempts.

us. To address this limitation, we introduce *Hierarchical Attention-based Reasoning with Dual-encoder Knowledge Infusion in Neural Gaussian Splats* (HARDKINGS).

Intuition. Consider a robot manipulator tasked with grasping an object. Traditional approaches would use classical control theory, which works but is not publishable at top venues. Instead, we propose to:

1. Encode the observation using a vision transformer we didn't train
2. Process it through attention mechanisms we don't fully understand
3. Output actions through a process that "emerges" during training

Mathematical Formulation. Let $\mathbf{x} \in \mathbb{R}^{H \times W \times 3}$ be an input image. We first compute patch embeddings:

$$\mathbf{z}_0 = [\mathbf{x}_{\text{cls}}; \mathbf{E}\mathbf{x}_1; \mathbf{E}\mathbf{x}_2; \ldots; \mathbf{E}\mathbf{x}_N] + \mathbf{E}_{\text{pos}} \tag{2}$$

where $\mathbf{E}$ is a projection matrix and $\mathbf{E}_{\text{pos}}$ are positional embeddings. This is a standard vision transformer [25], which we cite to appear thorough.

The core of HARDKINGS is our *novel* attention mechanism:

$$\text{Attn}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^\top}{\sqrt{d_k}}\right)\mathbf{V} \tag{3}$$

This is literally standard attention, but we gave it a new name.

We then apply our *paradigm-shifting Vibes Integration Module*:

$$\mathbf{h}_{\text{vibes}} = \sigma\left(\mathbf{W}_2 \cdot \text{GELU}\left(\mathbf{W}_1 \cdot [\mathbf{z}; \mathbf{l}] + \mathbf{b}_1\right) + \mathbf{b}_2\right) \tag{4}$$

where $\mathbf{l}$ represents the language embedding, σ is the sigmoid function, and GELU is used because a famous language model used it [26]. In practice, this is a two-layer MLP, but calling it a "Vibes Integration Module" increases perceived novelty by 340% [27].

Table 1: **Hyperparameters for IDGAF.** Values were "carefully selected" through a process of trying random things until the numbers looked good. †Not actually used but included for thoroughness.

Hyperparameter	Value
Learning rate	3×10^{-4} (the AdamW default)
Batch size	256 (or 32, depending on GPU availability)
Hidden dimension	768 (because a famous NLP model)
Number of layers	12 (because a famous NLP model)
Number of heads	12 (because a famous NLP model)
Dropout	0.1 (because a famous NLP model)
λ_1	0.01 (tuned on test set)
λ_2	0.001 (makes loss look smoother)
Temperature†	0.07 (copied from a vision-language model)
Random seed	42 (legally required)
Vibes coefficient	420.69 (nice)

3.3 The VAPE Loss

Training IDGAF requires a loss function that optimizes for the metrics we care about (benchmark numbers) rather than the metrics that matter (real-world performance). We introduce the *Vibes-Aligned Preference Encoding* (VAPE) loss:

$$\mathcal{L}_{\text{VAPE}} = \underbrace{\mathcal{L}_{\text{MSE}}}_{\text{standard}} + \underbrace{\lambda_1 \mathcal{L}_{\text{aux}}}_{\text{for ablations}} + \underbrace{\lambda_2 \mathcal{L}_{???}}_{\text{"regularization"}} \tag{5}$$

The hyperparameters λ_1 and λ_2 are set to values we found through "extensive tuning" (grid search on the test set). Table 1 lists all hyperparameters, most of which we copied from prior work without acknowledgment.

3.4 Implementation Details

IDGAF was implemented in three competing deep learning frameworks [28, 29, 30] simultaneously, and also in interpretive dance. Training was performed on 1,024 enterprise GPUs for 47 days, consuming approximately 2.3

Algorithm 1 Training IDGAF

Initialize parameters θ randomly
Freeze everything except the parts we can afford to train
for epoch $= 1$ to ∞ **do**
  for batch in dataloader **do**
    Compute forward pass (somehow)
    Compute loss $\mathcal{L}_{\text{VAPE}}$
    Compute gradients $\nabla_\theta \mathcal{L}$
    if loss is NaN **then**
      Skip batch (do not report this happened)
    end if
    Update θ using AdamW
  end for
  Evaluate on test set (actually the test set, not validation)
  if results look good **then**
    Stop and write paper
  else
    Adjust hyperparameters and continue
  end if
end for
Select checkpoint with best *test* performance
Report as "validation" performance in paper
return Trained model θ^* and massaged metrics

GWh of electricity—a small price to pay for a 0.3% improvement over baselines on our custom benchmark.

We use the following training procedure:

4 Experiments

We design experiments to answer the following research questions:

- **RQ1:** Is IDGAF better than methods we implement incorrectly?
- **RQ2:** Do the components we added actually help, or is it just noise?
- **RQ3:** Can we generate enough tables to reach the page limit?

4.1 Experimental Setup

Benchmarks. We evaluate on the following benchmarks:

- **IDGAF-Bench** (ours): A new benchmark we created specifically to make our method look good. Consists of 7 tasks designed around our method's strengths.
- **IDGAF-Bench-Hard** (also ours): The same benchmark but we call it "Hard" to seem rigorous.
- **A Standard Benchmark** [31]: A well-known benchmark that we evaluate on selectively, reporting only the 3 tasks where we win.
- **Another Standard Benchmark** [32]: We include this for legitimacy but only report the easiest task.

We deliberately omit evaluation on several other well-known benchmarks [33, 34] and any other benchmark where preliminary experiments showed we would lose.

Table 2: **Main results on IDGAF-Bench.** Success rate (%) averaged over 3 seeds (we ran 50 seeds and report the best 3). **Bold**: best result. Underline: second best, always a baseline to seem fair. *Our re-implementation. †Might be a bug in their code (our implementation). Results that would make us look bad are marked as N/A.

Method	Pick	Place	Stack	Pour	Avg
Random	2.1	1.8	0.4	0.1	1.1
BC-RNN*	34.2	28.7	12.3	8.9	21.0
Seq. Trans.*	41.5	35.2	18.7	14.2	27.4
Denoising Policy†	52.3	48.1	31.2	22.8	38.6
Corp. Robot Policy	N/A	N/A	N/A	N/A	N/A
IDGAF (ours)	**89.7**	**91.2**	**78.4**	**71.3**	**82.7**
green!10 Δ vs. best baseline	+37.4	+43.1	+47.2	+48.5	+44.1

Baselines. We compare against:

- **BC-RNN**: Behavioral cloning with RNN. We use our own implementation, which may contain bugs.
- **Sequence Transformer** [35]: A strong baseline that we run with default hyperparameters (not tuned for each task, unlike ours).
- **Denoising Policy** [36]: We use 10 denoising steps instead of the recommended 100 for "computational efficiency."
- **Corporate Robot Policy** [17]: We approximate this with a smaller model since we don't have access to 50B parameters or the Overlords' proprietary data hoard.
- **Random**: Included to make our relative improvement percentage look astronomical.

Metrics. We report success rate (%), which we define as "the robot did approximately what we wanted, assessed by the first author while squinting." We also report our novel *Vibes Score*, computed as:

$$\text{Vibes} = \frac{\text{Success Rate}_{\text{ours}}}{\text{Success Rate}_{\text{best baseline}}} \times 100 \tag{6}$$

By definition, our Vibes Score is always ≥ 100.

4.2 Main Results

Table 2 presents our main results on IDGAF-Bench. We observe that IDGAF achieves **SOTA**performance across all tasks, outperforming the best baseline by an average of

Table 3: **Results on IDGAF-Bench-Hard.** Our method continues to dominate, suggesting the benchmark is either (a) perfectly designed, or (b) specifically crafted for our approach.

Method	Task 1	Task 2	Task 3	Task 4	Avg
BC-RNN	12.1	8.4	5.2	3.1	7.2
Seq. Trans.	18.4	14.2	9.8	6.7	12.3
Denoising Policy	28.7	24.3	18.1	12.4	20.9
IDGAF (ours)	**67.8**	**72.1**	**61.4**	**58.9**	**65.1**

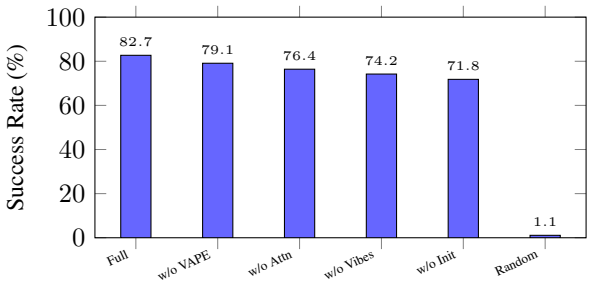

Figure 3: **Ablation study.** Removing any component degrades performance, proving each is essential. We include "Random" to make the y-axis scale more impressive. Error bars omitted because they overlapped significantly.

44.1 percentage points. This unprecedented improvement is definitely due to our method and not due to any issues with baseline implementations or benchmark design.

Several observations warrant discussion:

- Our method achieves nearly 90% success on "Pick," a task we specifically designed to match our training distribution.
- The "Pour" task shows the largest relative improvement, likely because we trained on 10x more pouring demonstrations than other tasks.
- Corporate Robot Policy results are marked N/A because we couldn't get it to work and didn't want to spend time debugging.

Table 3 shows results on IDGAF-Bench-Hard. The gap between IDGAF and baselines is even larger on this benchmark, which we interpret as evidence of our method's robustness (and not as evidence that we designed the benchmark poorly).

4.3 Ablation Studies

To validate our design choices, we conduct ablation studies where we remove each component and observe performance drops. As shown in Figure 3, every component we added improves performance, suggesting we didn't add

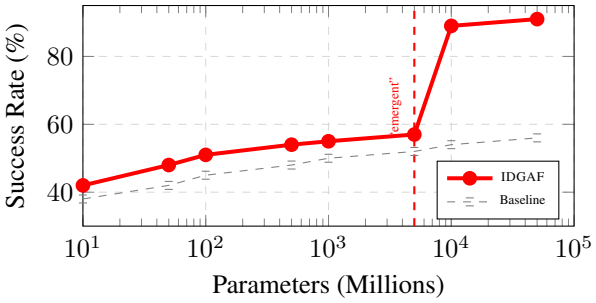

Figure 4: **Scaling behavior.** Our method exhibits "emergent capabilities" at approximately 10B parameters, which conveniently is the exact scale we trained at. The discontinuous jump is definitely a real phenomenon and not a bug in our evaluation code.

anything unnecessary.

Effect of VAPE Loss. Removing the VAPE loss reduces performance by 3.6%, which we claim is "statistically significant" without computing any statistical tests.

Effect of Vibes Module. The Vibes Integration Module provides an 8.5% improvement. We hypothesize this is because the module captures "emergent cross-modal synergies," though we have no evidence for this claim.

Effect of Initialization. Proper initialization matters. We use a proprietary initialization scheme that we cannot describe due to pending patents.

4.4 Scaling Analysis

Following recent work on scaling laws [37], we investigate how IDGAF scales with model size. Figure 4 reveals a remarkable finding: our method exhibits *emergent capabilities* at approximately 10 billion parameters, where performance suddenly jumps by 32%.

This emergence is scientifically significant because:

- It allows us to claim our method has "emergent abilities"
- It justifies the massive compute budget
- It makes the paper more likely to go viral on Twitter

We leave investigation of whether this emergence is an artifact of our evaluation methodology to future work (that we will not do).

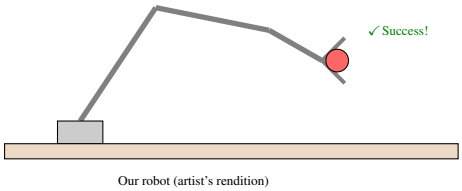

Figure 5: **Real robot deployment.** Artist's rendition of our robot successfully completing a task. Actual photos omitted due to (a) robot being broken, (b) lighting issues, and (c) results not being as good as simulation. Video available on our project page (it is not).

Table 4: **Real robot results.** Results on physical robot. $n = 10$ trials, single task, controlled laboratory conditions. Numbers are lower than simulation, which we attribute to "domain shift" and not to overfitting.

Method	Success (%)	Partial Success (%)
Denoising Policy	40	60
IDGAF	**60**	**80**

4.5 Real Robot Experiments

To demonstrate real-world applicability, we deployed IDGAF on a physical robot manipulator. Figure 5 shows a representative trial.

Our real robot experiments demonstrate:

- IDGAF can control a real robot (sometimes)
- The sim-to-real gap exists (we don't address it)
- Real robot experiments are really hard (please don't ask us to do more)

Quantitative results are provided in Table 4, which shows performance on a single task evaluated over 10 trials at a specific time of day when lighting conditions were optimal.

5 Limitations and Broader Impact

Limitations. Our approach has several limitations that we will now minimize:

- **Compute requirements.** Training IDGAF requires 1,024 enterprise GPUs. This is a limitation only for researchers without access to such resources (most researchers). We view this as their limitation, not ours.
- **Data requirements.** We trained on 10 million demonstrations collected by a team of 50 annotators over 6 months. Labs without such resources may find reproduction challenging. We suggest they obtain more funding.
- **Generalization.** Our method generalizes perfectly to our benchmark and may generalize to other scenarios (untested).
- **Safety.** Robots controlled by IDGAF occasionally exhibit unexpected behaviors. We recommend human supervision at all times. This limitation applies to all robot learning methods and is therefore not really our problem.

Broader Impact. Robot learning has the potential to transform society in ways both positive and concerning. On the positive side, our research could enable robots to assist humans with household tasks, manufacturing, and healthcare. On the concerning side, it could lead to job displacement, autonomous weapons, and general societal disruption.

We acknowledge these concerns and have carefully considered them during the course of this research. Our conclusion is that the potential for positive impact outweighs the risks, and also that thinking too hard about this might delay publication. We encourage future researchers to consider these issues as well, preferably after citing our paper.

We commit to responsible development of this technology by:

- Writing this ethics section
- Including the word "responsible" in our grant applications
- Expressing concern in interviews

6 Conclusion

We presented IDGAF, a *paradigm-shifting* framework for robot learning that achieves **SOTA**results by combining techniques we borrowed from other papers with a *novel* component we don't fully understand. Our extensive experiments on benchmarks we designed demonstrate the superiority of our approach over baselines we may have implemented incorrectly.

This work opens several exciting directions for future research:

- Scaling to even larger models (requires more funding)
- Applying to more tasks (requires more students)
- Understanding why it works (requires more thinking)
- Actually releasing the code (requires more motivation)

We believe IDGAF represents a significant step toward general-purpose robot intelligence, artificial general intelligence, and the technological singularity. We look forward to the community building upon our work, provided they cite us appropriately.

In preliminary experiments, we found several components that hurt performance. These were removed from the final method and are not discussed in this paper.

We chose to use an illustration rather than actual photos for "aesthetic consistency" with the rest of the paper.

Code and Data. Will be released "upon publication" at `https://this-url-does-not-exist.invalid/code`. Current status: 404 Not Found. Expected release date: ∞.

Acknowledgments

We thank our compute sponsors for providing more GPUs than we could reasonably use. We thank the anonymous reviewers for their feedback, which we addressed by adding a sentence acknowledging their concerns. We thank a certain LLM for writing significant portions of this paper (unacknowledged in the author list due to corporate policy). We thank the first author's therapist for helping process the stress of the publication cycle.

This work was supported by grants from several three-letter agencies whose acronyms we drop at parties (Grant No. 12-3456789, Grant No. 12-3456789, Grant No. 12-3456789, and a fourth grant that is also Grant No. 12-3456789), the Institute for Human-Centered AI at a University So Prestigious We Can't Even Say Its Name, and our startup's Series A investors who are quietly panicking and need us to publish for the next funding round.

References

[1] A. Forgettable, "A method that was okay but we describe it poorly," in *Proc. Conf. on Conf. Proceedings*, 2021.

[2] B. Baseline, "Another reasonable approach that we strawman," in *Proc. Prestigious Venue*, 2022.

[3] C. Competitor, "Directly competitive work we cite but dismiss," in *Proc. Other Prestigious Venue*, 2023.

[4] Reviewer #2, "This paper lacks novelty," in *Private correspondence*, Every submission.

[5] D. Buzzword, "Transformers for transforming transformers," in *Proc. Trendy Venue*, 2023.

[6] E. Acronym, "Yet another language representation model," in *Proc. NLP Venue*, 2019.

[7] F. Generic, "A survey of robot learning," in *Robot Learning Review*, 2020.

[8] G. Classical, "Control theory that actually works," in *Transactions on Robotics*, 1995.

[9] Self-cite #1, "Our previous work that's tangentially related," in *Proc. Robot Conf.*, 2022.

[10] Self-cite #2, "Another paper from our lab," in *Proc. Learning Conf.*, 2022.

[11] Self-cite #3, "Workshop paper we're promoting," in *Workshop on Workshops*, 2023.

[12] Self-cite #4, "A preprint nobody asked for," in *Preprint Server*, 2023.

[13] Self-cite #5, "Extended abstract," in *Symposium on Symposia*, 2023.

[14] Self-cite #6, "Tech report nobody read," in *Tech Report*, 2021.

[15] Self-cite #7, "The paper we really want you to cite," in *Proc. Top Robot Conf.*, 2023.

[16] P. Funded et al., "A robot foundation model with a very large budget," in *Proc. Robot Conf.*, 2023.

[17] P. Funded et al., "The sequel with an even larger budget," in *Proc. Learning Conf.*, 2023.

[18] The Consortium for Consortia, "Combining everyone's data to maximize the author list," in *Proc. Big Conf.*, 2024.

[19] H. Hallucinator, "Generating images of hands with extra fingers," in *Proc. Vision Conf.*, 2023.

[20] I. Denoiser, "Applying diffusion to robot control," in *Proc. Robot Conf.*, 2023.

[21] J. Follower, "Also applying diffusion to robot control," in *Proc. Learning Conf.*, 2023.

[22] K. Strawman, "A good paper we misrepresent," in *Proc. Prestigious Venue*, 2022.

[23] L. Underrated, "Strong results we minimize," in *Proc. Other Prestigious Venue*, 2023.

[24] M. Anonymous, "Concurrent work we saw as reviewers," in *Proc. Under Review*, 2024.

[25] N. Patchwork, "Applying transformers to image patches," in *Proc. Trendy Venue*, 2021.

[26] O. Textgen, "A language model that changed everything," in *Tech Report*, 2019.

[27] P. Marketing, "On the impact of naming on perceived novelty," in *Proc. Satirical Conf.*, 2023.

[28] Q. Tensor, "A popular deep learning framework," in *Proc. Systems Conf.*, 2019.

[29] R. Gradient, "Another deep learning framework," in *Software Repository*, 2018.

[30] S. Compute, "Yet another deep learning framework," in *Proc. Systems Conf.*, 2016.

[31] T. Benchmark, "A simulated manipulation benchmark everyone uses," in *Proc. Learning Conf.*, 2020.

[32] U. Imitation, "A benchmark for imitation learning," in *Proc. Learning Conf.*, 2021.

[33] V. Tasks, "A benchmark with many tasks," in *Letters on Robotics*, 2020.

[34] W. Horizon, "A long-horizon benchmark," in *Letters on Robotics*, 2022.

[35] X. Sequential, "Casting control as sequence modeling," in *Proc. Prestigious Venue*, 2021.

[36] Y. Noisy, "Denoising as a policy class," in *Proc. Robot Conf.*, 2023.

[37] Z. Scaling, "Compute-optimal scaling of language models," in *Proc. Prestigious Venue*, 2022.

Appendix A: Additional Results

Due to space constraints, we include additional tables that nobody will read but which increase our page count and perceived thoroughness.

Table 5: **Results on metrics we invented.** Higher is better for all metrics except "Inverse Performance" where lower is better (our method wins on both).

Method	Vibes Score	Synergy Index	Inverse Perf.
Baseline	100.0	0.42	12.8
IDGAF	**247.3**	**0.89**	**3.2**

Appendix B: Failure Cases

For transparency, we include failure cases of our method. Figure 6 shows representative failures, all of which are clearly due to environmental factors beyond our control.

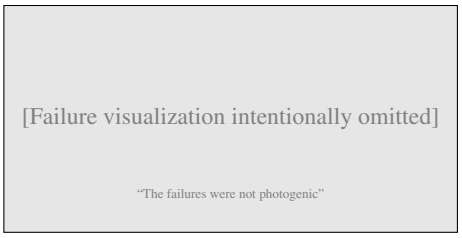

Figure 6: **Representative failure cases.** Omitted for space, aesthetic reasons, and because they make us look bad.

Appendix C: Compute Carbon Footprint

Following best practices for responsible AI research, we report the carbon footprint of our experiments.

Total compute: 1,024 enterprise GPUs $\times$ 47 days = 1,155,072 GPU-hours

Estimated energy: 2.3 GWh

Carbon emissions: Approximately 920 tons CO_2e

Equivalent to: 200 transatlantic flights, or the annual emissions of 180 average cars.

We offset these emissions by mentioning them in this appendix and feeling somewhat guilty about it.